# Translational Use Cases

Terry Braun

TrAPSS System as an Example of Translational Use Cases
(Candidate Disease Gene Prioritization and Mutation Screening)

**Use case 1: How we use genomic annotation tools now**

A researcher has 'case' and 'control' populations, where the 'case' is a particular disease phenotype of interest, and is interested in screening candidate disease genes (such as in a particular genome interval or locus from a linkage analysis, or candidate disease genes from an association study, or related genes through expression, function, or other methods for candidate nomination). The researcher has individuals, DNA, and candidate genes for mutation screening. (A researcher plans to screen individuals for sequence variants and demonstrate deleterious mutations).

Using a Genome browser, lab technicians may manually acquire data necessary to perform assays on the candidate genes. This includes that acquisition of the gene structure (gene sequence in a genomic context with introns, exons, and flanking genomic sequence). This data is used to generate assay primers for screening genes. Other sequence-based data may also be of use, such as: alternative transcripts, high similarity to other gene family members, known regulatory elements, existence and numbers of mRNAs and ESTs, etc. Additional non-sequence data may include expression data in the form of tissue specific expression, functional information and descriptions, pathway data, structural data, and links to various associated data (ex. primary literature).

Relevant issues for Architecture:

1. Completion of data models for genome annotation (for consumers).
2. API's to genomic annotation sources.

**Use case 2: How we plan to use genomic annotation tools**

A researcher plans to screen individuals for sequence variants and demonstrate deleterious mutations.

The researcher loads a manually currated list of candidate disease genes. The TrAPSS system queries various genome annotation resources (Ensembl, UCSC, NCBI) to acquire the necessary data elements. This includes flanking genomic sequence, exon sequence, intron sequence, exon-intron boundaries, UTR-exon boundaries, transcription start/stop, and translation start/stop.

With access to the sequence, automation of the generation of screening assays (SSCP or sequencing primers) is performed with primer3. Primers are reviewed, analyzed for quality (BLAT'ed against the genome), and stored. Primers are then ordered (electronically, in bulk) from a company that makes oligos (ex. Integrated DNA Technologies). The screening assay is performed, and results (gel shifts, sequence

variations) and images (SSCP, chromatograms) are recorded and linked with the data (genes, primers, plates of DNA).

Plates of DNA (containing 'cases' and 'controls' of individuals) are entered and are referenced by the screening results.

Relevant issues for Architecture:

1. Availability of generic interfaces to genomic annotation.
2. Tracking of data changes – genome reassemblies, etc.
3. Installation of supporting analytical components. For example, primer3 is required to generate primers. A BLAT server is necessary for quality analysis of primers.
4. Storage of patient information anonymously.
5. Security of patient information.

**Use case 3: How we hope to use genome annotation**

The TrAPSS system queries various genome annotation resources to infer regions of the candidate genes that are more likely to contain sequence variations that cause the disease phenotype. Analytical techniques examine conservation of conserved functional domains, conservation of genomic sequence across species, and other sequence-based elements (regulatory elements, repeats, GC-content, existence of previous mutations, SNPs, transmembrane domains, secondary structures, etc.). Users may provide guidance to the analytical techniques for importance of annotation features. Finally, additional non-sequence-based annotations may be used in conjunction with sequence-based features to provide an overall ranking of candidate genes. Non-sequence-based features may include expression data, pathway data, functional data, comparative genomic, linkage data, etc.

Relevant issues for Architecture:

1. Automatic/manual discovery of available annotation.